Encoding of High Dynamic Range Video With a Model of Human Cones

Hateren, J.H. van

Published in:
Acm transactions on graphics

Publication date:
2006

Encoding of High Dynamic Range Video With a Model of Human Cones

J. H. VAN HATEREN
University of Groningen and Netherlands Institute for Neuroscience

A recently developed quantitative model describing the dynamical response characteristics of primate cones is used for rendering high dynamic range (HDR) video. The model provides range compression, as well as luminance-dependent noise suppression. The steady-state (static) version of the model provides a global tone mapping algorithm for rendering HDR images. Both the static and dynamic cone models can be inverted, enabling expansion of the HDR images and video that were compressed with the cone model.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image Generation—*Display algorithms*; I.4.2 [**Image Processing and Computer Vision**]: Compression (Coding)—*Approximate methods*

General Terms: Algorithms

Additional Key Words and Phrases: High dynamic range, HDR video, tone mapping, adaptation, video processing, visual perception, human cones

1. INTRODUCTION

Visual scenes in the real world often contain a wide range of luminances. Neither conventional display equipment nor printing technologies allow a direct representation of the typical luminance range needed. Therefore, the images and movies acquired from scenes are usually transformed by a nonlinear, compressive operation to reduce the range of luminances needed for display. For rendering still images, such techniques are generally known as tone mapping algorithms [Devlin 2002; Reinhard et al. 2005]. An important goal of rendering an image is to keep the result as close as possible to the original, with "close" loosely defined as similar in terms of human perception [Pattanaik et al. 1998]. Not surprisingly, known facts on human perception, including quantitative models, play an important role in guiding the search for good rendering algorithms [Ferwerda et al. 1996; Pattanaik et al. 2000; Mantiuk et al. 2004; Krawczyk et al. 2005].

Apart from predicting human perception, there is a second reason why knowledge of the human visual system can help to suggest good algorithms for rendering. This is because visual systems are faced with a similar problem as imaging systems: The neurons in the brain, including the cone photoreceptor cells in the retina, have a rather limited dynamic range, in the order of 10^2. Scenes, on the other hand, contain luminance ranges of typically 10^3 for medium-contrast scenes, and up to 10^4–10^6

Authors' address: J. H. van Hateren, Department of Neurobiophysics, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands; email: j.h.van.hateren@rug.nl.

for scenes containing, for example, both sunlit areas and deep shadows, specular highlights, or simultaneous views of indoor and outdoor scenes. Cones contain a range of mechanisms that are tailored to reduce large luminance ranges. Such mechanisms may be, and have been, used as inspiration for designing tone mapping algorithms to accomplish range reduction [Ledda et al. 2004; Reinhard and Devlin 2005]. The present article follows this approach, and extends it to the time domain. This has now become possible by the recent development of a quantitative model of primate cones and horizontal cells [van Hateren 2005] which adequately describes measurements performed on cells in the macaque retina [Smith et al. 2001; Lee et al. 2003]. The macaque retina is generally considered to be functioning in a nearly identical way to the human retina, and in the following, the model for macaque cones will be assumed to apply to human cones equally well.

The cone model describes the fast adaptation performed by the enzymatic amplifier inside the cone that transduces light into electrical signals. The slower process of pigment bleaching, reducing the effective input to cones in bright daylight conditions, is not included in the model. However, it is fairly straightforward to add pigment bleaching as a front-end module [van Hateren and Snippe (in preparation)]. Adaptive shifts of the tone mapping curve that are analogous to the action of pigment bleaching have been used extensively for rendering both static high dynamic range (HDR) images [Baxter et al. 1982; Normann et al. 1983; Daly 1993; Tumblin et al. 1999; Ashikmin 2002; Reinhard and Devlin 2005] and dynamic HDR video [Pattanaik et al. 2000; Durand and Dorsey 2000; Scheel et al. 2000; Artusi et al. 2003; Bennett and McMillan 2005; Irawan et al. 2005].

The cone model presented here is particularly well-suited for rendering HDR video, where it performs two major functions. First, it reduces the dynamic range needed for display by a combination of dynamic nonlinearities. Second, it reduces noise by low-pass filtering that, as part of the dynamics of the model, automatically adapts to prevailing scene luminances. In addition, the steady-state (static) version of the model can be used for rendering HDR images. This version is fully invertible, and the dynamic version of the model is also invertible, apart from the low-pass filtering. Therefore, the images (and video) with compressed dynamic range can be expanded again to obtain images identical to (and video quite similar to) the originals. This particular combination of compressing and expanding is also known as companding (e.g., Li et al. [2005]), and the algorithms are well-suited for companding either HDR images or HDR video.

2. ENCODING BY HUMAN CONES

There is adaptation and gain control at nearly every stage in the human visual system, and even luminance gain control is not fully confined to the cones. To provide some perspective on what is and is not modeled here, I will first give a schematic overview of the processing steps that are thought to occur.

2.1 Dynamic Range Control at the Early Stages of the Human Visual System

Figure 1 shows a minimal scheme of early visual processing. The scene is imaged by the cornea and lens onto the retina, with the pupil controlling both the image quality (a large pupil implies poor image quality because of lens aberrations) and retinal illuminance (proportional to pupil area). In studies on retinal processing, it is customary to use a unit called a "troland" (td) as the measure of retinal illuminance. The scene luminance in cd/m^2 multiplied by the pupil area in mm^2 yields the retinal illuminance in td. The model presented next predicts the output of the cones as a function of input retinal illuminance (in td), and therefore does not include variations in pupil diameter as one of the ways used by the visual system to control sensitivity. However, for daylight conditions, the variation in pupil diameter is limited [Westheimer 1986], and not expected to provide much more than a factor of two or three for controlling retinal illuminance. Moreover, the dynamical behavior of the pupil is

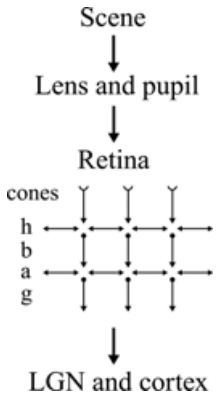

Fig. 1. Simplified scheme of the first stages of visual processing in the human visual system. Light projected onto the retina is transduced into electrical signals by the cones, which connect to horizontal cells (h) and bipolar cells (b). The latter connect to amacrine cells (a) and ganglion cells (g), which project to the LGN (lateral geniculate nucleus), where the signals are relayed to the cortex.

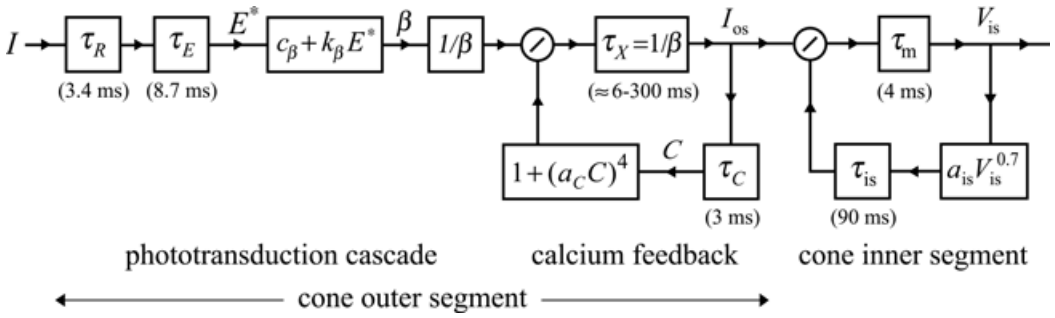

Fig. 2. Model describing the dynamical response of cones. Light I is filtered by a series of first-order low-pass filters (τ), and transformed by several nonlinearities. The dynamics are dominated by two divisive feedback loops. Parameters and variables as determined in van Hateren [2005]: The retinal illuminance I is expressed in trolands (= luminance times pupil area in mm^2), $\tau_R = 3.4$ ms, $\tau_E = 8.7$ ms, $c_\beta = 2.8{\cdot}10^{-3}$ (ms)$^{-1}$, $k_\beta = 1.6{\cdot}10^{-4}$ (ms)$^{-1}$/td, $\tau_C = 3$ ms, $a_C = 9{\cdot}10^{-2}$, $\tau_\mathrm{m} = 4$ ms, $a_\mathrm{is} = 7{\cdot}10^{-2}$, $\tau_\mathrm{is} = 90$ ms. The variable I_os is implicitly multiplied by a unitary dimensional conversion constant to make it dimensionless; C and V_is are dimensionless as well in the model as presented here.

complex, partly depending on cognitive factors. It is also relatively slow (with a time constant in the order of a second) compared with the dominating dynamic processes in the cones (with time constants considerably shorter than one second, see Figures 2 and 3). To avoid such complexities, and to provide a convenient conversion factor between td and cd/m^2, we will assume a fixed pupil with an area of 10 mm^2 (diameter 3.6 mm). This pupil diameter occurs at daylight luminances between 10 and 100 cd/m^2 [Westheimer 1986]. The result is that 1 cd/m^2 is taken here as equivalent to 10 td (see also the top and bottom axes of Figure 5(a)).

Light projected onto the retina is absorbed by visual pigment in the cones and rods. Because perception at daylight intensities is dominated by cones, the rods will not be considered here. The cones come in three varieties with slightly different visual pigments: S-cones (short wavelength-sensitive, in the blue part of the spectrum), M-cones (medium wavelength-sensitive, in the green-to-yellow part of the spectrum), and L-cones (long wavelength-sensitive, in the green-to-yellow-to-red part of the spectrum).

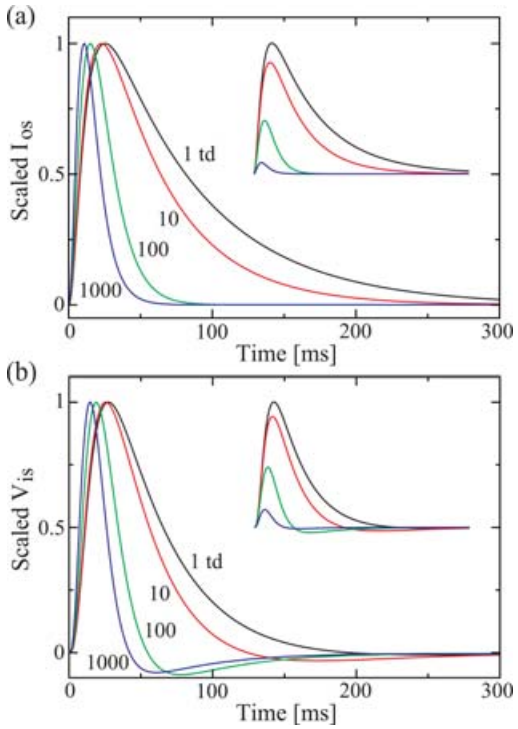

Fig. 3. Pulse responses of I_{os} (a) and V_{is} (b) to short (1 ms) light pulses of 100 td superimposed on background illuminances of 1,10,100, and 1,000 td. The main curves are inverted and normalized and the insets show the responses before normalization.

At high luminances (retinal illuminances larger than several thousand troland) a significant part of the visual pigment becomes bleached (inactivated), resulting in a proportional reduction of cone sensitivity. Bleached pigment is regenerated with a time constant in the order of a minute, thus sensitivity control by pigment bleaching is much slower than the dominating dynamical processes in the cones.

The cones contain machinery for transducing the incoming light into electrical signals, and also contain powerful gain control mechanisms for reducing the dynamic range of the signals (detailed in the next section). The cones both transduce the light and adapt to it, strictly localized to their own field of view [MacLeod et al. 1992; Lee et al. 1999]. However, they interact with a layer of neurons, the horizontal cells ("h" in Figure 1, with two subtypes), which provide lateral connectivity (i.e., with cones that look into different parts of space and that may have a different spectral sensitivity). The horizontal cells are believed to further shape the output signals of the cones based on the luminance averaged over short stretches of time and over small spatial neighborhoods. The result of this operation is a suppression of low spatial and temporal frequencies (i.e., sharpening in space and time), and possibly some further gain control. However, the details of how this aspect functions in the human retina are not yet clear.

The signals from cones, as modified by horizontal cells, are subsequently transferred to, first, the bipolar cells ("b" in Figure 1, with several subtypes) and then to the ganglion cells ("g" in Figure 1, including subtypes specializing in fast, achromatic processing, or slower, chromatic processing; see, e.g., van Hateren et al. [2002]). A second site of lateral connectivity is provided by the amacrine cells ("a" in Figure 1), operating at the level where bipolar cells feed their signals into the ganglion cells.

There are many subtypes of amacrine cells, of which the function in the human retina is mostly not well-known. The ganglion cells are the output neurons of the retina, and finally send their signals to higher areas in the brain (via the LGN to the cortex).

Several processes in the retina are relevant for dynamic range reduction. First, the cones provide luminance gain control. Suppression of low spatial and temporal frequencies by horizontal cells further reduces the dynamic range needed for the signals represented in the bipolar cells. At the level of ganglion cells, chromatic opponency (roughly speaking, red-green, i.e., L/M, and blue-yellow, i.e., S/(L + M) opponency) and contrast gain control further reduce the dynamic range needed for the signals in various subtypes of ganglion cells. Several of these processes, which partly consist of slowly adapting components, presumably contribute to luminance gain control as well. Nevertheless, the bulk of luminance gain control is provided by cones.

2.2 A Model of Human Cones

There has been tremendous progress in the past 10–20 years in unraveling the molecular steps leading from the absorption of light to electrical responses in rods and cones (reviews: Pugh and Lamb [2000]; Fain et al. [2001]). What is known about the dynamics of these processes was recently translated into a comprehensive system model [van Hateren 2005] that successfully describes a recent series of measurements on horizontal cells in the macaque retina [Smith et al. 2001; Lee et al. 2003], as well as measurements on human cones using the electroretinogram [van Hateren and Lamb 2006]. The model was validated for the illuminance range covered by these measurements, approximately 0.5–2000 td. The part of the model that describes the cones is shown in Figure 2. Note that this model describes the response characteristics of a single cone, which receives light from its own unique patch in visual space. All components in the model therefore represent purely temporal, not spatial, filters. When light I (expressed in troland) is absorbed by visual pigment, it starts a cascade of enzymatic amplifiers (for details, see van Hateren [2005]), which can be represented by two first-order temporal low-pass filters (τ_R and τ_E) and a gain (k_β). A first-order low-pass filter with a time constant τ, transforming an input $x(t)$ into an output $y(t)$, can be described by the following differential equation:

$$\frac{dy}{dt} + \frac{1}{\tau} y = \frac{1}{\tau} x. \tag{1}$$

The coefficient in front of x is chosen here such that the DC-gain of the filter equals one: when $\frac{dy}{dt} = 0$, we find $y = x$. The term $k_\beta E^*$ in Figure 2 represents the light-dependent activity of an enzyme, which has a residual activity, c_β, even in complete darkness. This enzyme breaks down the substance (cGMP) that controls the current across the cell membrane (I_{os}). The breakdown of cGMP can be described by a nonlinear differential equation [Nikonov et al. 2000; van Hateren 2005], which has two major consequences. First, it produces a strong nonlinearity ($1/\beta$ in Figure 2, with $\beta = c_\beta + k_\beta E^*$), contributing to dynamic range compression. Second, it affects temporal processing according to a first-order low-pass filter (τ_X) that is dependent on the retinal illuminance ($\tau_X = 1/\beta$). Because τ_X depends on β, which itself depends on the retinal illuminance (through E^*), τ_X can be considered to adapt, continuously and automatically, to the prevailing light level. The low-pass filtering is slower when luminance is low (i.e., τ_X is approximately 300 ms for 1 td, cf. 6 ms for 1000 td) and thus provides extended integration time when the signal-to-noise ratio (SNR) is also small due to photon noise (see the following). The breakdown of cGMP is counteracted by the production of cGMP regulated by a highly nonlinear feedback loop under the control of intracellular calcium. The behavior of this system is well-understood (see van Hateren [2005], in particular, the supplementary material). The feedback loop moderates the effect of the strong $1/\beta$ nonlinearity, and it also counteracts, to some extent, the large luminance-dependent variation of the time constant τ_X. The output is a current across the cell membrane, I_{os}, generated by the outer

segment of the cone (i.e., the cone- or rod-shaped part of the cone containing the visual pigment and phototransduction machinery). Figure 3(a) shows the effect of various levels of background illumination (1, 10, 100, and 1,000 td) on the response of I_{os} to a short (1 ms) pulse of 100 td superimposed on the background light. The main curves are inverted and normalized in order to facilitate the comparison of temporal dynamics, and the inset shows the responses before normalization. The reduction in gain with increasing illuminance (an effect producing range compression) is caused by the combined action of the $1/\beta$ nonlinearity and the calcium feedback loop (Figure 2). Figure 3(a) clearly shows that I_{os} is considerably slower at low backgrounds than at high backgrounds, that is, the integration time of the cone depends on the background intensity.

The current I_{os} subsequently flows into the cone's inner segment, and is further modified by a feedback loop describing the nonlinear current-voltage relationship of the membrane of the inner segment. The resulting signal, represented by the membrane potential of the cone, V_{is}, is then transmitted to horizontal and bipolar cells. The feedback loop representing the cone's inner segment contains a relatively slow low-pass filter (τ_{is}), leading to biphasic impulse responses of the cell. Figure 3(b) shows the responses of V_{is} to the same backgrounds and light pulses as used for Figure 3(a). Again, the main curves are inverted and normalized, and the inset shows the responses before normalization. Although the curves can be interpreted as impulse responses, it should be realized that the model has nonlinear dynamics. The exact shape of the curves, such as the size of rebounds at (b), therefore depends not only on the background light level, but also on the size of the pulse.

The feedback loop representing the inner segment contributes to dynamic range compression because the biphasic impulse responses of Figure 3(b) suppress the DC-component of the illuminance. However, its contribution to range compression is considerably less than that attributable to the processing in the cone's outer segment [van Hateren 2005]. Moreover, I have found that including the inner segment is less suitable for rendering HDR video. The reason is that it produces visible artifacts (slow over- and under-shoots) that trail moving high-contrast edges, caused by the biphasic impulse response (Figure 3(b)). The remainder of this article will therefore concentrate on the outer segment (I_{os}), which does not produce such visible artifacts when used for the purpose of rendering. However, purely for the purpose of companding (compressing and expanding), it would be advantageous and straightforward to include the inner segment loop as well, both in the forward and inverse models.

2.3 Rendering and Companding

Figure 4 again shows on the left the algorithm proposed here for compressing and rendering. Because of the $1/\beta$ operation, I_{os} is large for small luminances, and small for large luminances. Electrical responses of rods and cones are thus inverted when compared with the light level: increasing light levels produce decreasing I_{os}. For the purpose of rendering, I_{os} must therefore be inverted and scaled. The extremes of I_{os} follow from $I = 0$ and $I = \infty$, with

$$I_{os,min} = I_{os}|_{I=\infty} = 0, \tag{2}$$

and $I_{os,max}$ the solution of

$$I_{os,max} = I_{os}|_{I=0} = \frac{1/c_\beta}{1 + (a_C I_{os,max})^4}, \tag{3}$$

where we used the fact that all low-pass filters in the model have unity DC-gain, and thus can be omitted for the steady-state solution. Equation (3) can be readily solved numerically. The inverted and

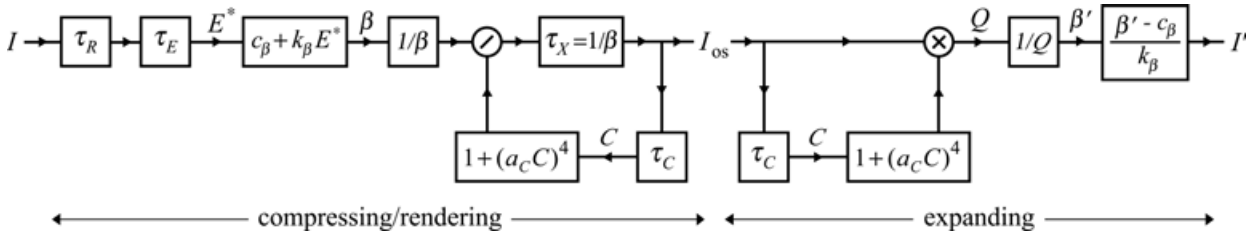

Fig. 4. Cone model up to the generation of $I_{\rm os}$, and expansion stage producing reconstruction I' of the retinal illuminance I. Parameters as in Figure 2.

scaled output, o, then follows from

$$o = r\,\frac{I_{\rm os,max} - I_{\rm os}}{I_{\rm os,max} - I_{\rm os,min}} = r\,(1 - I_{\rm os}/I_{\rm os,max}), \tag{4}$$

with r the required output range.

The right of Figure 4 shows the expansion. All low-pass filters that are on the direct, forward path of the compression stage have been omitted from the expansion stage. The rationale for this omission is threefold. First, an inverted low-pass filter will, unless regularized, strongly enhance the high-frequency noise that may have been added to $I_{\rm os}$ (e.g., because of limited numerical precision, discretization of $I_{\rm os}$, or an additional lossy encoding and decoding of $I_{\rm os}$). Second, the forward low-pass filters in the cone (τ_R,τ_E, and τ_X) are thought to be instrumental for reducing noise in the signal (partly photon noise, partly noise arising from other sources), serving the purpose of increasing the SNR of the resulting signal that is passed on to horizontal and bipolar cells. It makes sense, therefore, to conserve this supposedly beneficial effect, and not undo it by high-pass filtering. Finally, omitting inverted low-pass filters makes the expansion particularly simple and fast. For frame rates with interframe times considerably longer than the time constant τ_C (= 3 ms), that is, for most regular video, the filter τ_C can be neglected. The expansion stage then reduces to a simple static nonlinearity, identical to Eq. (8), which follows.

Whereas expansion is fast, compression requires a significant amount of computation because of the nonlinear feedback loop. I implemented the algorithm by using a modification of the exponential Euler integration scheme for each low-pass filter [Brown 2000; van Hateren 2005]. The scheme is in fact an autoregressive moving average filter with only three nonzero coefficients: for a time step Δt, the output $y(t)$ of a first-order low-pass filter with time constant τ in response to an input $x(t)$ is given, at times $t = n\Delta t$, as

$$y_n = f_1 y_{n-1} + f_2 x_{n-1} + f_3 x_n, \tag{5}$$

with

$$\begin{aligned} f_1 &= \exp(-1/\tau') \\ f_2 &= \tau' - (1+\tau')\exp(-1/\tau') \\ f_3 &= 1 - \tau' + \tau'\exp(-1/\tau') \\ \tau' &= \tau/\Delta t. \end{aligned} \tag{6}$$

In contrast to the basic forward Euler scheme (with $f_1 = 1 - 1/\tau'$, $f_2 = 0$, and $f_3 = 1/\tau'$), this scheme is stable for any value of τ', and it is more accurate than the regular exponential Euler scheme (with $f_1 = \exp(-1/\tau')$, $f_2 = 0$, and $f_3 = 1 - \exp(-1/\tau')$; this scheme is often used in neurophysiological simulations, which are dominated by first-order dynamics). Because of nonlinear feedback, time step Δt for the feedback loop still needs to be small. For the calculations in this article, $\Delta t =$ 0.1 ms. This

value is somewhat smaller than is strictly necessary, but when it is significantly increased, the feedback can become unstable, giving rise to spurious oscillations. Compressing a 128 × 128 pixel movie at an output frame rate of 50 frames per second required a computation of approximately 1.4 seconds per frame on a 3 GHz Pentium 4. The supplementary material of van Hateren [2005] provides an example source code of an extended version of the present model.

Although beyond the scope of the present article, it is likely that the compression algorithm can be made considerably faster. The essence of the algorithm is a filter (with impulse responses as in Figure 3) of which the gain and temporal bandwidth is continuously adjusted to the input illuminance (actually, a low-pass filtered and offset version of the input, namely, β in Figure 2). Rather than implicitly calculating the gain and temporal dynamics with the full model for each video, it may be possible to store the gain and appropriate time constants in look-up tables, and to use these values to continuously adjust simplified (higher-order) low-pass filters that approximate the dynamics of the full model.

The expansion is always fast, even when the filter τ_C is not neglected. Because the expansion stage contains no feedback, only feedforward, the modified exponential Euler scheme only needs to be computed with a time step equal to the interframe time. For movies with reasonable frame rates and sizes, this can be easily performed in real time by current hardware.

2.4 Steady-State Compression and Expansion Curves

For rendering and companding HDR images, the model can be used in its steady-state version. The compression stage follows from solving I_{os} from

$$I_{\mathrm{os}} = \frac{1/(c_\beta + k_\beta I)}{1 + (a_C I_{\mathrm{os}})^4}, \tag{7}$$

which can be readily done numerically. Note that this need only to be done once because I_{os} is a fixed function of I . Figure 5(a) shows (inverted and normalized) I_{os} as a function of I (in td). For convenience, the ordinate scale is also shown as scene luminance (cd/m^2, assuming a 10 mm^2 pupil). As can be seen, the tone mapping curve extends over a luminance range of approximately 10^4. The curve has a sigmoid shape similar to curves that are commonly used for rendering HDR images. An example of such a widely used curve is the nonlinearity, $L^n/(L^n + \sigma^n)$, with L luminance and n and σ parameters determining the detailed shape of the curve [Baxter et al. 1982; Normann et al. 1983; Daly 1993; Pattanaik et al. 2000]. For comparison, the dashed curve in Figure 5(a) shows the best fit with this function, where the fit yielded $\sigma = 235$ td and $n = 0.76$. This value for the coefficient n is similar to values commonly used.

The inverse operation (expansion) can be expressed in closed form, and follows from the right of Figure 4:

$$I' = \left[\frac{1}{I_{\mathrm{os}}(1 + (a_C I_{\mathrm{os}})^4)} - c_\beta \right] / k_\beta. \tag{8}$$

Figure 5(b) shows the reconstructed scene luminance I' as a function of (inverted and normalized) I_{os}.

3. RESULTS

For testing the cone algorithm, I used several HDR images and movies. I will concentrate on three different examples to follow, but results on other images (e.g., Figure 7) and movies are similar. The HDR radiance maps used for Figures 6 and 8 and for the movies are of Belgium House (courtesy of Dani Lischinski, provided at http://www.cs.huji.ac.il/ ∼danix/hdr/pages/belgium.html) and Stanford Memorial Church (courtesy of Paul Debevec, provided at http://www.debevec.org/Research/HDR/). Furthermore, I constructed HDR video by two different methods. For the first method, I used HDR static

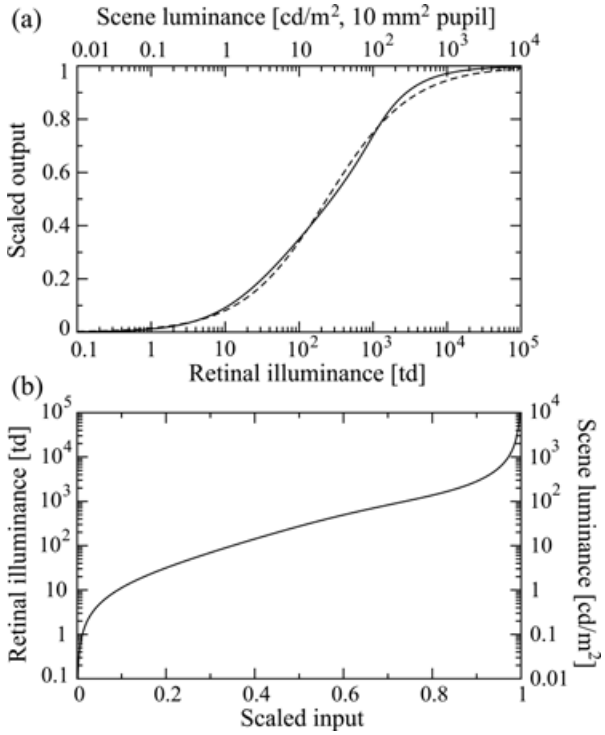

Fig. 5. Steady-state tone mapping (compression) curve ((a), continuous line) and its inverse, the expansion curve (b). The compression curve at (a) was computed by numerically solving I_{os} from Eq. (7) by using routine rtbis of Numerical Recipes [Press et al. 1992] for finding the root of the lefthand-side minus the righthand-side of Eq. (7). The dashed line at (a) shows a fit with $L^n/(L^n + \sigma^n)$, see text for details.

images. A small window sliding over such an image provides frames for a movie, in effect simulating a camera panning the scene. For the second I generated the movie "RNL" (Rendering with Natural Light). It consists of moving virtual objects illuminated by a natural scene [Debevec 1998], and was constructed using Radiance lighting simulation software [Ward Larson 1994] and source files made available by Paul Debevec at http://athens.ict.usc.edu/RNL/Source. Whereas the first type of movie only contains global motion, the second type also contains local motion of objects relative to the background.

3.1 Rendering and Reconstructing HDR Images

The cone model requires image intensities expressed in trolands, a measure of retinal illuminance that does not specify how the light energy is distributed across the spectrum. However, there are three spectral types of cones, namely, S, M and L, and some way to define trolands for each of these types is therefore needed for using the cone model. In retinal research it is customary to define so-called 'blue trolands' for the S-cones (with equivalent definitions for the M- and L-cones), where 1 blue troland is the light level that excites the S-cone equally strongly as an equal-energy white of 1 td (equal energy means that the light energy is distributed evenly over the spectrum; see also Boynton and Kambe [1980]). In principle, a linear transformation is needed to convert RGB pixel values in the image into L-, M- and S-cone excitation values because the spectral properties of R-, G, and B-channels will, in general, not match the spectral sensitivities of L-, M-, and S-cones, respectively. Because the main goal in this article is not to predict exactly what all types of cones would do in response to a particular

Fig. 6. Original images, and images rendered with the come algorithm.

image, but rather to present an algorithm that is useful for rendering, we will simplify matters, and let each R-, G-, and B-channel respond as if it each represents a cone type. An equal-energy white of 1 td should then correspond to the value 1 in the R-channel (i.e., 1 'R-troland'), and similarly to values of 1 in the G- and B-channels. A complication here is that although HDR images provide a linear intensity scale, the absolute calibration of this scale onto trolands is usually not exactly known. Therefore, I scaled the HDR images by using the log-average, a measure often used to determine the typical luminance of HDR images. The log-averaged illuminance is defined here as $\bar{I} = \exp(\langle \ln I_R \rangle/3 + \langle \ln I_G \rangle/3 + \langle \ln I_B \rangle/3)$, with brackets denoting averaging over all pixel values of a particular color channel (e.g., I_R for the R-channel) in the image. Scaling is then defined by assuming, for a particular image, a specific value for $\bar{I}$, and subsequently scaling all pixels to produce this $\bar{I}$. Note that scaling defined in this way guarantees that equal values in R-, G-, and B-channels (presumably representing a neutral gray) will still be equal (and therefore still produce gray) after nonlinear transformation by the cone model.

Using the static compression curve of Figure 5(a), based on Eq. (7), HDR images can thus be rendered with $\bar{I}$ as the only adjustable parameter. By default, I am assuming $\bar{I} = 200$ td because this is close to the (logarithmic) mid-point of the dynamic range of the tone mapping curve in Figure 5(a). Figure 6 shows results for two HDR images and a still frame from the RNL movie. The examples show the original images rendered with $\gamma = 2.2$ (see Section 3.2 to follow for a discussion of gamma correction). For cone rendering, I found that the results were best without gamma correction; with gamma correction, the contrast of the images became too low (see Section 3.2). Whereas $\bar{I} = 200$ td appears to be a reasonable choice for rendering Belgium House (top right) and RNL (bottom right), Memorial Church gives too much impression of an outdoor, daylight scene when rendered at $\bar{I} = 200$ td (Figure 6, bottom row). This can be corrected by rendering at a lower value of $\bar{I}$, with $\bar{I} = 50$ td shown in the figure. Because the cone rendering is limited to a total dynamic range of approximately 10^4 (Figure 5), and both the Belgium and Memorial images encompass a total range in the order of 10^6, the cone rendering has some saturation in the darkest and brightest parts of these images. Reconstructions of the original images for all examples in Figure 6, using Eq. (8), are, as expected, indistinguishable from the originals, and are therefore not reproduced in the figure. However, it should be realized that all computations were performed in floating point, and discretization issues were not explored here (see also the Discussion and Conclusions section). Figure 7 provides a few more examples of cone-rendered HDR images.

3.2 Cone Rendering Does not Require Gamma Correction

Because of the physics of cathode ray tubes, display equipment is generally assumed to have a nonlinear relationship between digital pixel values I_{comp} in the computer and the resulting pixel luminance I_{disp} on the display [Poynton 1996]. To first approximation, this can be described as $I_{\text{disp}} = I_{\text{comp}}^{\gamma}$ with γ =2.5. It is therefore customary to compensate a measured or calculated scene luminance I_{scene} as $I_{\text{comp}} = I_{\text{scene}}^{1/\gamma}$. This is called gamma correction. In practice, a slight undercompensation is often used, with γ =2.2. This increases the contrast of the resulting image, giving a better subjective picture quality under the conditions of a dim viewing environment [Poynton 1996].

Although gamma correction can be viewed as a linearization of the input-output chain of imaging, its main usefulness has to do with properties of the human visual system [Poynton 1996]. This system codes primarily relative steps in luminance, not absolute steps, with a resolution in the order of 1%. Whereas a step from, for example, 20 to 21 is well-perceived, a step from 200 to 201 is not. It would therefore be inefficient to map the (typically, only 256) levels of I_{comp} linearly onto the luminance levels of the display. At low levels of I_{comp}, there would be visible discretization and noise artifacts in I_{disp}, whereas the high relative luminance resolution of I_{disp} at high levels of I_{comp} would not be visible. The gamma of the display equipment solves this problem to a large extent by compressing low levels of I_{comp} and stretching high levels of I_{comp} when transforming to I_{disp}. In effect, this makes much better use of the available number of distinct levels of I_{comp} by (approximately) equalizing the perceptual differences between subsequent levels (see also Ward Larson et al. [1997] and Irawan et al. [2005] for an extension of this idea).

This equalization is most effective when all levels of I_{comp} occur with equal probability because the coding is then most efficient [Shannon 1948]. However, the distribution of luminances occurring in real scenes is far from even. Figure 8 (top row) shows the probability density function (pdf) of the scenes of Figure 6. The pdfs are quite skewed, with long tails, which is in fact typical for real scenes [Richards 1982; Laughlin 1983; van Hateren and van der Schaaf 1996; van der Schaaf 1998, p. 57 and 58). Figure 8 shows illuminances between 0 and $10\bar{I}$, but higher illuminances are not shown (ignoring the high end of the tail, i.e., 8.3% of the pixels in Belgium House, 6.4% in Memorial Church, and 3.9% in RNL).

The effect of a gamma correction with $\gamma = 2.2$ is shown in the middle row of Figure 8 (same range and percentages of ignored pixels as in the top row). The result is that the range of I_{comp} is filled more evenly.

Fig. 7. Examples of images rendered with the cone algorithm (upper left) office (published in Durand and Dorsey [2002]), courtesy of Byong Mok Oh, rendered with $\bar{I} = 200$ td (upper right) tree, courtesy of OpenEXR, $\bar{I} = 200$ td (lower left) UH Library, courtesy of Sean O'Malley, $\bar{I} = 50$ td (lower right) bottles, courtesy of Sean O'Malley, $\bar{I} = 200$ td.

Interestingly, a qualitatively similar result is produced by the static cone model (Figure 8, bottom row). The pdf is now (with the exception of the $\bar{I} = 50$ td rendering of Memorial Church) even more equalized. There is no long tail at high luminances, and all pixels are included in the pdf shown.

From the aforementioned analysis, we can conclude that the static cone model performs an operation which is qualitatively (but not quantitatively) similar to a gamma correction: It transforms the skewed pdf of scene luminances into a well-equalized pdf of I_{comp}, well-suited to be presented, through $I_{\text{disp}} = I_{\text{comp}}^{\gamma}$, to the human eye. Performing a gamma correction on top of the cone transformation would therefore be counterproductive: The resulting pdf would have a tail to the left, and a large fraction of the pixels would be mapped onto high luminances on the display, where the (absolute) discriminability of the human eye is low.

3.3 Rendering and Reconstructing HDR Video

Video recorded in real time always contains noise, not only instrumental noise, but also photon noise. Photon noise cannot be avoided because there is only a limited number of photons that can be captured by each pixel per shutter time (with shutter time usually taken as equal to or smaller than interframe

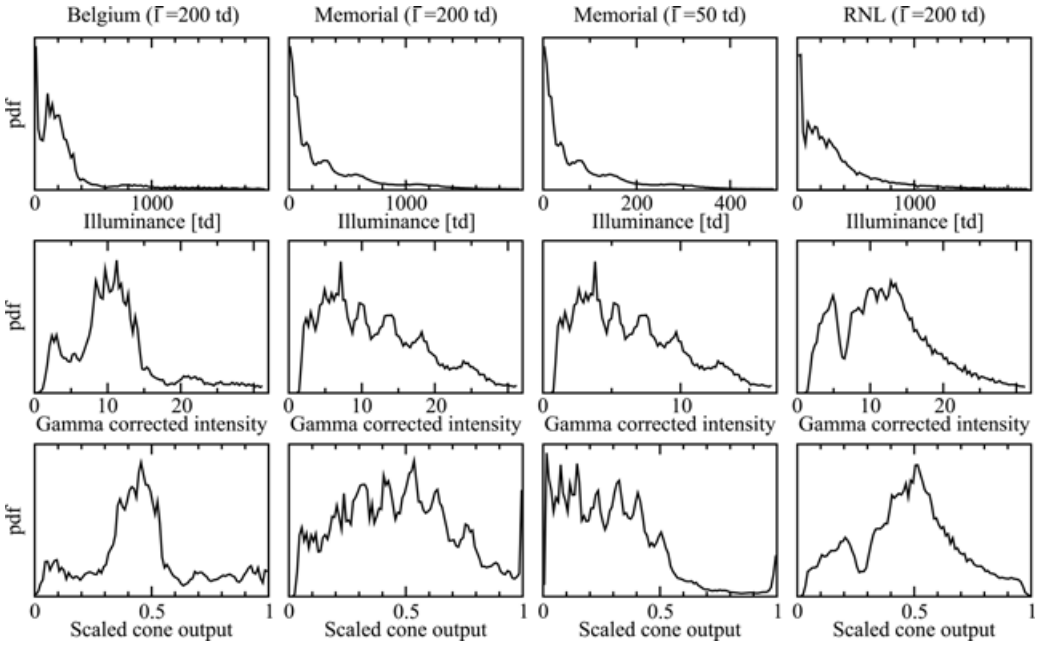

Fig. 8. Probability density functions of scene illuminances (top row: the linear illuminance scale runs from 0 to 10 $\bar{I}$), of gamma corrected illuminances for the same illuminance range (middle row, $\gamma = 2.2$), and of the inverted and scaled output of the cone algorithm (bottom row, full range).

time). The amount of photons captured depends, among other things, on scene luminance, the aperture of the camera lens, pixel size, and shutter time. Whereas the signal is proportional to the number of captured photons, noise is proportional to the square root of the number of captured photons because photon noise has the properties of Poisson noise. The SNR therefore decreases with decreasing scene luminance. Moreover, instrumental noise usually also increases at low luminances because more amplification is needed. One of the attractive aspects of the cone model is that it adjusts its low-pass filtering depending on the input luminance, and thus adapts to the expected SNR. To simulate and test the effect of this on real video, I added Poisson noise to the constructed videos described earlier. The video was scaled to trolands similarly as for the HDR images, with $\bar{I}$ now defined with averaging over all pixels in the movie. The equivalent number of photons captured by a pixel and the noise then follow from a conversion factor, which is taken here as 1 td = 400 photons/second/pixel for the movies belgium.mpg and rnl.mpg (both with $\bar{I} = 200$ td). This conversion corresponds approximately to a camera with 6×6 μm^2 pixels and an F8 aperture (following the analysis by Delbrück and Mascarenhas [1997]). For the calculation, it was further assumed that shutter time was equal to interframe time. For the movie memorial.mpg, the noise was, for the purpose of presentation, made more visible by assuming 1 td = 50 photons/second/pixel for the $\bar{I} = 50$ td rendering. This would correspond to a camera with smaller pixels or a higher F-number than assumed before, or a shutter time much smaller than the interframe time. This level of noise may be similar to what is actually present in the cones of the human retina with this scene and luminance because 1 td corresponding to 50 absorbed photons/second/cone is a fair estimate of the order of magnitude [Smith and Lamb 1997; Schneeweis and Schnapf 1999]. This higher level of photon noise compared with (high-end) video cameras can be attributed primarily to the fact that the light collecting area of cones (typically, in the order of

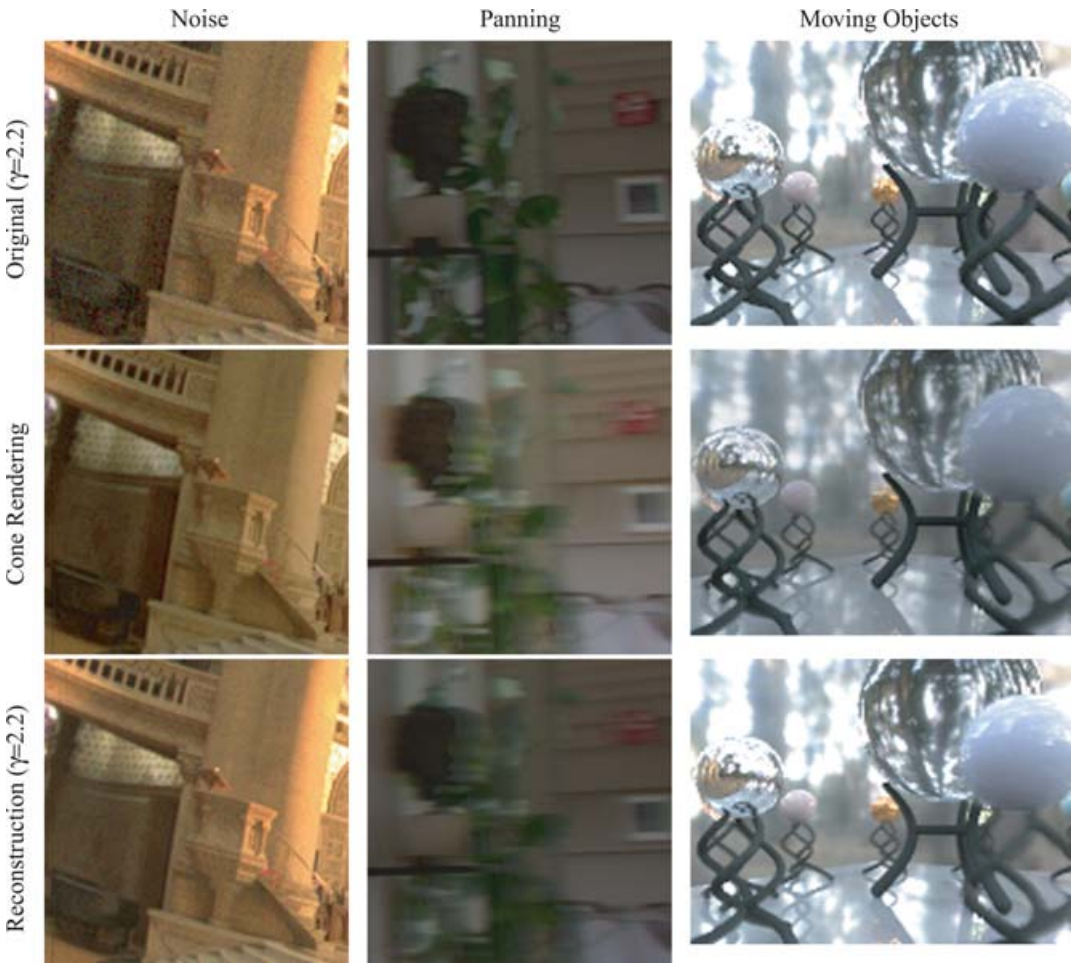

Fig. 9. Frames from the movies memorial.mpg (column "Noise", 128 × 128 pixel frames, scaling $\bar{I} = 50$ td, noise level determined by assuming 1 td = 50 photons/second/pixel), belgium.mpg (column "Panning", 128 × 128 pixel frames, scaling $\bar{I} = 200$ td, 1 td = 400 photons/second/pixel), and rnl.mpg (column "Moving Objects", 160 × 120 pixel frames, scaling $\bar{I} = 200$ td, 1 td = 400 photons/second/pixel). The first two movies were constructed by extracting frames from the HDR images Memorial Church and Belgium House, the latter movie was constructed using Radiance, with the "Rendering with Natural Light" source files. The upper row shows original frames rendered with $\gamma = 2.2$, the middle row the output of the cone model, and the lower row the reconstruction of the original from the cone output ($\gamma = 2.2$). To facilitate the comparison, the cone latency was compensated by showing the original frames two frames earlier (40 ms) than the frames of the cone rendering and reconstruction.

$2 \times 2\ \mu\text{m}^2$ in the fovea, larger further away from the fovea) is smaller than the typical pixel area of such cameras.

The resulting movies are accompanying this article (also available at http://hlab.phys.rug.nl/demos/cone_rendering), with representative frames shown in Figure 9. The column "Noise" illustrates the capacity of the cone algorithm to attenuate noise. It shows frames corresponding to memorial.mpg at a steady camera position with the original rendered at $\gamma = 2.2$, the cone rendering according to the left of Figure 4, and the reconstruction rendered at $\gamma = 2.2$ again. The reconstruction is nearly identical to the original, although close scrutiny of the frames reveals that the noise level in the reconstruction

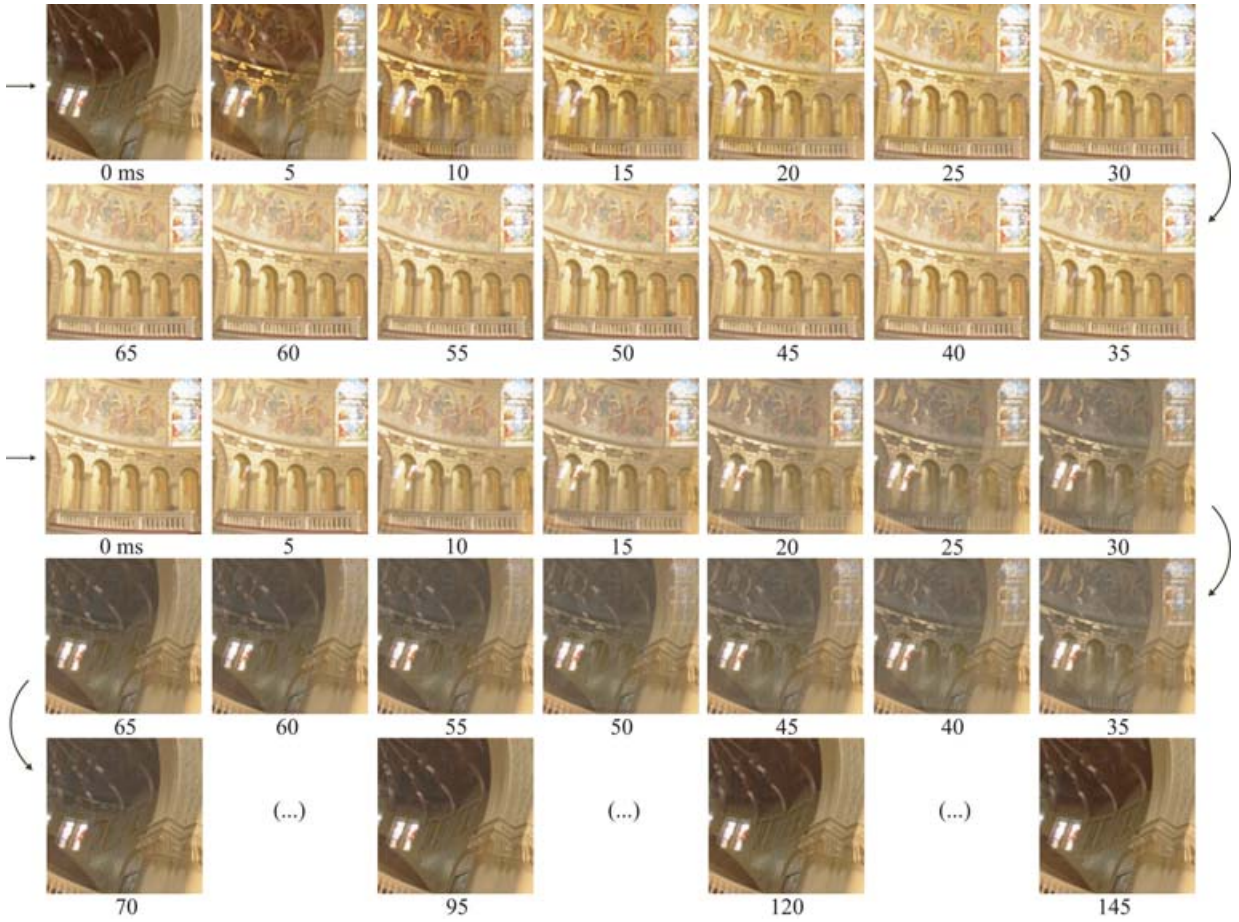

Fig. 10. Frames from the movie adap.mpg.

(and in the cone rendering) is lower than that in the original (see also memorial.mpg). Also note that the range is compressed in the cone rendering compared with the original and reconstruction (see the bright area on the far right, and also other areas in Figure 6). Finally, note that the cone rendering is less yellowish and reddish than the original and reconstruction (see again Figure 6). This color change is caused by the fact that the gain in each of the three color channels is regulated independently by the cone algorithm, which results in an elementary (von Kries-type; see Wyszecki and Stiles [1982]) form of color constancy.

The noise reduction comes at a price, namely, motion blur. The column "Panning" (from belgium.mpg) shows the effect of a moving camera on the cone rendering and reconstruction. Similarly, the column "Moving Objects" (frames from rnl.mpg) shows the effect of different motion speeds: For example, the blue sphere and its stand in the foreground are slightly blurred in the cone rendering and reconstruction, whereas there is little blur for the large central sphere and its stand, which are moving only slowly in this section of the movie. Also note that with respect to range compression, the background has more structure in the cone rendering than in the original and reconstructed scenes.

How the adaptation of gain and integration time in the cone develops can be seen in Figure 10 (from the movie adap.mpg). Here, two scene cuts are simulated, one from a rather dark to a rather bright

scene, and the other from bright to dark. In order to show the dynamics more clearly, frames are shown every 5 ms. The upper sequence shows that the new, bright scene quickly displaces the old, dark scene. However, note that the afterimage of the windows in the lower left corner of the original scene only becomes fully imperceptible after about 50 ms. The lower sequence shows that it takes more time for the dark scene to replace the bright one. Traces of the arches and rim above the arches remain visible until 70–95 ms after the scene cut. The reason for this asymmetry in adaptation is the time constant τ_X in Figure 2: When a bright pixel appears, E^* and β become large, and hence $\tau_X = 1/\beta$ small. The scene therefore establishes itself quickly at I_{os}. When a dark pixel appears, β becomes small, and hence τ_X large. It then takes more time to reach the steady-state value of I_{os}. The advantage is that dark, noisy parts of the scene are subjected to increased integration times (Figure 3(a)), which reduces noise. Note, however, that even for dark scenes, the entire adaptation process will be finished within a few frames if the frame rate is modest (e.g., 20 ms between frames).

4. DISCUSSION AND CONCLUSIONS

In this article, a dynamical model of the human cone was applied to rendering HDR video and HDR images. The model not only reduces the dynamic range needed for representing the video and images, but also reduces the noise in HDR video in a luminance-dependent way by continuously adjusting the cone integration time (Figures 3 and 10). The original images and video can be reconstructed, lossless for images and lossy for video, using the inverse model of the cone.

The model of the cone contains mainly fast processes, and can therefore quickly adapt to video input. It can handle an input range of up to 10^4, which appears to be consistent with the limited luminance range humans can simultaneously perceive in a single scene. The model does not include the range of additional, mostly slower, mechanisms that the human visual system uses for shifting the compression curve (Figure 5(a)) along the luminance axis, with varying pupil diameter, pigment bleaching, and adaptive processes elsewhere in the retina. At still lower luminances than considered here, τ_R and k_β possibly increase, and there is, of course, the shift from cone- to rod-dominated vision.

The main component that could be added to the model, extending its use to a wider range of (bright) luminance levels, is a front-end module equivalent to pigment bleaching (and also incorporating some of the effects of variable pupil diameter). Pigment bleaching starts to become important for illuminances larger than a few thousand troland, and is essentially equivalent to adding a gray filter ("dark glasses") in front of the cones. It reduces, in a linear way, the effective retinal illuminance that should go into the cone model. Processes that mimic pigment bleaching have been used extensively in the literature on HDR rendering (e.g., Pattanaik et al. [2000], Irawan et al. [2005]. The basic scheme is to let a low-pass filtered version of the pixel luminance continuously shift the position of the tone mapping curve along the luminance axis. In the present scheme, this would amount to a variable multiplication factor between scene luminance (in cd/m^2) and the effective input (in td) to the cone model of Figure 4. Although pigment bleaching in cones represents a rather slow low-pass filter (with a time constant in the order of a minute), a time constant in the order of a few seconds may be more appropriate for rendering HDR video as a compromise between realism and perceptibility.

The static compression curve (Figure 5(a)) represents a global tone mapping algorithm: It is the same for all pixels in an image. The performance of the algorithm for HDR images is therefore not fundamentally different from other global algorithms. The perceived contrast is inevitably small in high-luminance parts of the image because the human visual system (mostly the cones) has the property of contrast constancy: Equal contrasts lead to similar response amplitudes in the cones, independently of the luminance [van Hateren 2005]. Global tone mapping algorithms also provide some form of contrast constancy (by using a monotonic, compressive nonlinearity), which is thus added to that of the perceiving human visual system. The result is too much compression, producing a decreasing perceived contrast

with increasing luminance. Only local tone mapping algorithms that vary in space can circumvent this problem (see, e.g., Fattal et al. [2002]). Nevertheless, the algorithm presented here gives reasonable results, and is convenient because it contains no free parameters for images for which the original scene luminance is known, and only a single free parameter, $\bar{I}$, when the calibration is not known. However, the static algorithm is not advocated here as preferable to existing global algorithms; a detailed comparison with the performance of other algorithms is beyond the scope of this article. The static algorithm is presented here primarily as an aid in understanding the performance of cones in dynamical circumstances.

For HDR video, the cone provides not only range compression, but also noise reduction (see Bennett and McMillan [2005] for an alternative approach). Again, there are no free, adjustable parameters when the calibration is known. The algorithm will automatically give an output similar to that of human cones for luminances in the range for which the cone model was validated (approximately 0.5–2000 td, see van Hateren [2005]). This is an important range because it covers common levels of interior lighting and also the lower range of outdoor light levels. When outdoor light levels are in the range that produces pigment bleaching, the input gain could be simply shifted according to the aforementioned scheme.

For rendering the example movies, the cone model was used with the same time constants as the human cone. Similar to the problem with range compression (performed twice, first by the model and then by the viewer's cones looking at the output of the model), there is also a problem with motion blur. Blur produced by actual cones is added to the blur produced by the cone model. Usually, the resulting blur will be only increased by approximately $\sqrt{2}$ compared with the expected blur (as would have occurred when watching the original scene), and may be hardly noticeable. However, for movie scenes where the viewer's eye can track moving objects within the scene, the blur produced by the model for such objects can be significantly larger than expected. In such cases, it may be better to use the cone model with all time constants reduced by a fixed factor. This would not pose a problem for noise reduction because of the presumably lower levels of photon noise in camera pixels compared with human cones. An alternative solution would be to add motion compensation to the model.

The algorithms can be used not only for rendering, but also for the purpose of companding by using the compression and expansion stages of Figure 4. A full analysis of this application is beyond the scope of this article (but see the application of companding presented in Li et al. [2005] for the kind of approach meant here). Preliminary observations indicate that discretizing I_{os} to 6 bits per color channel before performing the expansion is sufficient for producing reconstructed images and video that are perceptually quite similar to the originals. In these tests, the algorithm itself (both compression and expansion) were done, as before, in floating-point arithmetic, and only the intermediate stage, I_{os}, was discretized. However, if the high dynamic range information must be retained, I have found that I_{os} has to be discretized to approximately 10 bits per color channel. This suggests that this algorithm may be similarly effective (even without explicit separation of color and luminance) as various standard formats for storing HDR images, which typically use 32 bits or more for the three color channels combined (http://www.anyhere.com/gward/hdrenc/). In these standard formats, the color information is usually first separated from the luminance information, and the latter is subsequently strongly compressed. However, this is the reverse order of what the retina does: first, a strong compression of each color channel (mainly performed by the cones), and only later an approximate separation of color and luminance (particularly observable at the level of ganglion cells).

Finally, the adaptive low-pass filtering that the cone model performs on HDR video produces an I_{os} that is smoother in time and contains lower noise levels than the original movie. Both of these properties are advantageous for encoding I_{os} as, for example, an mpeg movie. Rendering such a movie would then imply mpeg decoding followed by the expansion stage shown at the right in Figure 4.

MOVIE CAPTIONS

memorial.mpg. Frames were obtained by shifting a 128 × 128 pixel frame over the image Memorial Church (courtesy of Paul Debevec), generating 250 frames/second at a shift of 1 pixel/frame. The input to the cone model was provided at 250 frames per second, holding the current frame as input during the interframe period. The output of the model was sampled at 50 frames/second. The movie shows on the left the original ($\gamma = 2.2$), in the middle the cone rendering, and on the right the reconstruction ($\gamma = 2.2$), all at 50 frames/second. For this movie, the 250 frames/second input was reduced to 50 frames/second by block-averaging consecutive groups of 5 frames. The scaling was $\bar{I} = 50$ td; photon shot noise consistent with 1 td = 50 photons/second/pixel was added to each input frame. The movie is shown at twice the original size for the sake of clarity.

belgium.mpg. Frames were obtained from the image Belgium House (Fattal et al. [2002]). The scaling was $\bar{I} = 200$ td; 1 td = 400 photons/second/pixel. Further details are as for memorial.mpg.

rnl.mpg. Frames were obtained using the "Rendering with Natural Light" script [Debevec 1998], assuming 50 frames/second. The input to the model was provided at 50 frames per second, holding the current frame as input during the interframe period. The movie shows at the top left the original ($\gamma = 2.2$), at the bottom the cone rendering, and at the top right the reconstruction ($\gamma = 2.2$) at 50 frames/second, all at 1.5 times the original size. The scaling was $\bar{I} = 200$ td; 1 td = 400 photons/second/pixel.

adap.mpg. Frames were obtained from the image Memorial Church. The input to the cone model was provided at 200 frames per second, holding the current frame as input during the interframe period. The output of the model was sampled at 200 frames/second (5 ms between each frame), but rendered at 25 frames/second in the movie. The movie therefore runs 8 times slower than real time, and is intended to show the (fast) adaptive processes of the cone model. The movie shows on the left the original ($\gamma = 2.2$), and on the right the cone rendering, at twice the original size. The scaling was $\bar{I} = 200$ td; 1 td = 400 photons/second/pixel.

ACKNOWLEDGMENTS

I wish to thank H.P. Snippe for comments on this manuscript.

REFERENCES

ASHIKHMIN, M. 2002. A tone mapping algorithm for high contrast images. In *EGWR: Proceedings of the 13th Eurographics Workshop on Rendering*. 145–156.

ARTUSI, A., BITTNER, J., WIMMER, M., AND WILKIE, A. 2003. Delivering interactivity to complex tone mapping operators. In *EGSR: Proceedings of the 14th Eurographics Symposium on Rendering*. 38–44.

BAXTER, B. S., RAVINDRA, H., AND NORMANN, R. A. 1982. Changes in lesion detectability caused by light adaptation in retinal photoreceptors. *Investigative Radiology 17*, 394–401.

BENNETT, E. P. AND MCMILLAN, L. 2005. Video enhancement using per-pixel virtual exposures. *ACM Trans. Graph. 24*, 3, 845–852.

BOYNTON, R. M. AND KAMBE, N. 1980. Chromatic difference steps of moderate size measured along theoretically critical axes. *Color Res. Appl. 5*, 13–23.

BROWN, K. S. 2000. Lead-Lag algorithms. http://www.mathpages.com/home/kmath198/kmath198.htm.

DALY, S. 1993. The visible differences predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*. A.B. WATSON, Ed. MIT Press, Cambridge, MA. 179–206.

DEBEVEC, P. E. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Computer Graphics Annual Conference Series (SIGGRAPH)*. 189–198.

DELBRÜCK, T. AND MASCARENHAS, S. M. 1997. Notes on practical photometry. http://www.ini.unizh.ch/~tobi/anaprose/recep/practicalPhotometry.pdf.

DEVLIN, K. 2002. A review of tone reproduction techniques. Tech. Rep. CSTR-02-005, Department of Computer Science, University of Bristol.

DURAND, F. AND DORSEY, J. 2000. Interactive tone mapping. In *Proceedings of the Eurographics Workshop on Rendering Techniques*. 219–230.

DURAND, F. AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Computer Graphics Annual Conference Series (SIGGRAPH)*, 257–266.

FAIN, G. L., MATTHEWS, H. R., CORNWALL, M. C., AND KOUTALOS, Y. 2001. Adaptation in vertebrate photoreceptors. *Physiological Rev. 81*, 117–151.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Trans. Graph. 21*, 249–256.

FERWERDA, J. A., PATTANAIK, S. N., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. 249–258.

IRAWAN, P., FERWERDA, J. A., AND MARSCHNER, S. R. 2005. Perceptually-Based tone mapping of high dynamic range image streams. In *EGSR: Proceedings of the 16th Eurographics Symposium on Rendering*. 231–242.

KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Lightness perception in tone reproduction for high dynamic range images. *Comput. Graph. Forum 24*, 3, 635–645.

LAUGHLIN, S. B. 1983. Matching coding to scenes to enhance efficiency. In *Physical and Biological Processing of Images*. O. J. BRADDICK, and A. C. SLEIGH, Eds. Springer, New York. 42–52.

LEDDA, P., SANTOS, L. P., AND CHALMERS, A. 2004. A local model of eye adaptation for high dynamic range images. In *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, and Visualisation*. 151–160.

LEE, B. B., DACEY, D. M., SMITH, V. C., AND POKORNY, J. 1999. Horizontal cells reveal cone type-specific adaptation in primate retina. *Proceedings of the National Academy of Science U.S.A., 96*, 14611–14616.

LEE, B. B., DACEY, D. M., SMITH, V. C., AND POKORNY, J. 2003. Dynamics of sensitivity regulation in primate outer retina: The horizontal cell network. *J. Vision 3*, 513–526.

LI, Y., SHARAN, L., AND ADELSON, E. H. 2005. Compressing and companding high dynamic range images with subband architectures. In *ACM Trans. Graph. 24*, 3, 836–844.

MACLEOD, D. I. A., WILLIAMS, D. R., AND MAKOUS, W. 1992. A visual nonlinearity fed by single cones. *Vision Res.* 32, 347–363.

MANTIUK, R., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2004. Perception-Motivated high dynamic range video encoding. *ACM Trans. Graph. 23*, 731–741.

NIKONOV, S., LAMB, T. D., AND PUGH, E. N., JR. 2000. The role of steady phosphodiesterase activity in the kinetics and sensitivity of the light-adapted salamander rod photoresponse. *J. General Physiology 116*, 795–824.

NORMANN, R. A., BAXTER, B. S., RAVINDRA, H., AND ANDERTON, P. J. 1983. Photoreceptor contributions to contrast sensitivity: Applications in radiological diagnosis. *IEEE Trans. Syst. Man Cybernetics 13*, 944–953.

PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. 287–298.

PATTANAIK, S. N., TUMBLIN, J., YEE, H., AND GREENBERG, D. P. 2000. Time-Dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. 47–54.

POYNTON, C. 1996. Gamma. In *A Technical Introduction to Digital Video*. JohnWiley, New York. http://www.poynton.com/notes/TIDV/index.html.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., AND FLANNERY, B. P. 1992. *Numerical Recipes in Fortran*. Cambridge University Press, New York.

PUGH, E. N., JR. AND LAMB, T. D. 2000. Phototransduction in vertebrate rods and cones: Molecular mechanisms of amplification, recovery and light adaptation. In *Handbook of Biological Physics,* vol. 3. D. G. STAVENGA et al. Eds. Elsevier, Amsterdam. 183–254.

REINHARD, E. AND DEVLIN, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Trans. Visualization Comput. Graph. 11*, 13–24.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Rendering*. Morgan Kaufmann, San Francisco, CA.

RICHARDS, W. A. 1982. Lightness scale from image intensity distributions. *Appl. Optics 21*, 2569–2582.

SCHEEL, A., STAMMINGER, M., AND SEIDEL, H.-P. 2000. Tone reproduction for interactive walkthroughs. *Comput. Graph. Forum 19*, 3, 301–312.

SCHNEEWEIS, D. M., AND SCHNAPF, J. L. 1999. The photovoltage of macaque cone photoreceptors: Adaptation, noise, and kinetics. *J. Neurosci. 19*, 1203–1216.

SHANNON, D. E. 1948. The mathematical theory of communication. *Bell Syst. Tech. J. 27,* 3–91.

SMITH, N. P., AND LAMB, T. D. 1997. The a-wave of the human electroretinogram recorded with a minimally invasive technique. *Vision Res. 37*, 2943–2952.

SMITH, V. C., POKORNY, J., LEE, B. B., AND DACEY, D. M. 2001. Primate horizontal cell dynamics: An analysis of sensitivity regulation in the outer retina. *J. Neurophysiology. 85*, 545–558.

TUMBLIN, J., HODGINS, J. K., AND GUENTHER, B. K. 1999. Two methods for display of high contrast images. *ACM Trans. Graph. 18*, 56–94.

VAN DER SCHAAF, A. 1998. Natural image statistics and visual processing. PhD thesis, University of Groningen. http://irs.ub.rug.nl/ppu/166956252.

VAN HATEREN, J. H. 2005. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *J. Vision* 5, 331–347.

VAN HATEREN, J. H. AND LAMB, T. D. 2006. The photocurrent response of human cones is fast and monophasic. *BMC Neurosci. 7*, 34.

VAN HATEREN, J. H., RÜTTIGER, L., SUN, H., AND LEE B. B. 2002. Processing of natural temporal stimuli by macaque retinal ganglion cells. *J. Neurosci. 22*, 9945–9960.

VAN HATEREN, J. H. AND VAN DER SCHAAF, A. 1996. Temporal properties of natural scenes. In *Proceedings of the IS and T/SPIE Conference on Electronic Imaging: Science & Technology*. San Jose, CA. 139–143.

WARD LARSON, G. J. 1994. The RADIANCE lighting simulation and rendering system. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*. 459–472.

WARD LARSON, G. J., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. Visualization Comput. Graph. 3*, 291–306.

WESTHEIMER, G. 1986. The eye as an optical instrument. In *Handbook of Perception and Human Performance*. K. R. BOFF, et al., Eds. John Wiley, New York.

WYSZECKI, G. AND STILES, W. S. 1982. *Color Science*. John Wiley, New York.

Received June 2005; revised September 2005; accepted May 2006